

Claims

1. Apparatus for building a stochastic model of a data sequence, said data sequence comprising time related symbols selected from a finite symbol set, the apparatus comprising:

an input for receiving said data sequence,

a tree builder for expressing said symbols as a series of counters within nodes, each node having a counter for each symbol, each node having a position within said tree, said position expressing a symbol sequence and each counter indicating a number of its corresponding symbol which follows a symbol sequence of its respective node, and

a tree reducer for reducing said tree to an irreducible set of conditional probabilities of relationships between symbols in said input data sequence.

2. Apparatus according to claim 1, said tree reducer comprising a tree pruner for removing from said tree any node whose counter values are within a threshold distance of counter values of a preceding node in said tree.

3. Apparatus according to claim 2, wherein said threshold distance and tree construction parameters are user selectable.

4. The apparatus of claim 3, wherein said user selectable parameters further comprise a tree maximum depth.

5. The apparatus of claim 3, wherein said user selectable parameters further comprise an algorithm buffer size

6. The apparatus of claim 3, wherein said user selectable parameters further comprise values of pruning constants.

7. The apparatus of claim 3, wherein said user selectable parameters further comprise a length of input sequences.

8. The apparatus of claim 3 wherein, said user selectable parameters further comprise an order of input symbols.

9. Apparatus according to claim 2, wherein said tree reducer further comprises a path remover operable to remove any path within said tree that is a subset of another path within said tree.

10. Apparatus according to claim 1, wherein said sequential data is a string comprising consecutive symbols selected from a finite set.

11. The apparatus of claim 10, further comprising an input string permutation unit for carrying out permutations and reorganizations of the input string using external information about a process generating said string.

12. Apparatus according to claim 1, wherein said sequential data comprises output data of a manufacturing process.

13. Apparatus according to claim 12, wherein said output data comprises buffer level data.

14. Apparatus according to claim 12, said process comprising feedback.

15. Apparatus according to claim 1, wherein said sequential data comprises seismological data.

16. Apparatus according to claim 1, wherein said sequential data is an output of a medical sensor sensing bodily functions

17. Apparatus according to claim 16, wherein said output comprises visual image data and said medical sensor is a medical imaging device.

18. Apparatus according to claim 1, wherein said sequential data is data indicative of operation of cyclic operating machinery.

19. Apparatus for determining statistical consistency in time sequential data, the apparatus comprising

a sequence input for receiving sequential data,

a stochastic modeler for producing at least one stochastic model from at least part of said sequential data,

and a comparator for comparing said sequential stochastic model with a prestored model, thereby to determine whether there has been a statistical change in said data.

20. Apparatus according to claim 19, wherein said stochastic modeler comprises:

a tree builder for expressing said symbols as a series of counters within nodes, each node having a counter for each symbol, each node having a position within said tree, said position expressing a symbol sequence and each counter indicating a number of its corresponding symbol which follows a symbol sequence of its respective node, and

a tree reducer for reducing said tree to an irreducible set of conditional probabilities of relationships between symbols in said input data sequence.

21. Apparatus according to claim 19, said prestored model being a model constructed using another part of said time-sequential data.

22. Apparatus according to claim 19, said comparator comprising a statistical processor for determining a statistical distance between said stochastic model and said prestored model.

23. Apparatus according to claim 22, said statistical distance being a KL statistic.

24. Apparatus according to claim 22, said statistical distance being a relative complexity measure.

25. Apparatus according to claim 22, wherein said statistical distance comprises an SPRT statistic.

26. Apparatus according to claim 22, wherein said statistical distance comprises an MDL statistic.

27. Apparatus according to claim 22, wherein said statistical distance comprises a Multinomial goodness of fit statistic.

28. Apparatus according to claim 22, wherein said statistical distance comprises a Weinberger Statistic.

29. Apparatus according to claim 20, said tree reducer comprising a tree pruner for removing from said tree any node whose counter values are within a threshold distance of counter values of a preceding node in said tree.

30. Apparatus according to claim 29, wherein said threshold distance is user selectable.

31. The apparatus of claim 30, wherein user selectable parameters further comprise a tree maximum depth.

32. The apparatus of claim 30, wherein user selectable parameters further comprise an algorithm buffer size.

33. The apparatus of claim 30, wherein user selectable parameters further comprise values of pruning constants.

34. The apparatus of claim 30, wherein user selectable parameters further comprise a length of input sequences.

35. The apparatus of claim 30, wherein user selectable parameters further comprise an order of input symbols.

36. Apparatus according to claim 29, wherein said tree reducer further comprises a path remover operable to remove any path within said tree that is a subset of another path within said tree.

37. Apparatus according to claim 19, wherein said sequential data is a string comprising consecutive symbols selected from a finite set.

38. The apparatus of claim 37, further comprising an input string permutation unit for carrying out permutations and reorganizations of said sequential data using external information about a process generating said data.

39. Apparatus according to claim 19, wherein said sequential data comprises output data of a manufacturing process

40. Apparatus according to claim 39, said process comprising feedback.

41. Apparatus according to claim 19, wherein said sequential data comprises seismological data.

42. Apparatus according to claim 19, wherein said sequential data is an output of a medical sensor sensing bodily functions.

43. Apparatus according to claim 19, wherein said sequential data is data indicative of operation of cyclic operating machinery.

44. Apparatus according to claim 22, wherein said data sequence comprises indications of a process state, the apparatus further comprising a process analyzer for using said statistical distance measure as an indication of behavior of said process.

45. Apparatus according to claim 22, wherein said data sequence comprises indications of a process state, the apparatus further comprising a process controller for using said statistical distance measure as an indication of behavior of said process, thereby to control said process.

46. Apparatus according to claim 23, wherein said data sequence comprises multi-input single output data.

47. Apparatus according to claim 22, wherein said data sequence comprises financial behavior patterns.

48. Apparatus according to claim 22, wherein said data sequence comprises time sequential image data sequences said model being usable to determine a statistical distance therebetween.

49. Apparatus according to claim 48, said image data comprising medical imaging data, said statistical distance being indicative of deviations of said data from an expected norm.

50. Apparatus according to claim 22, applicable to a database to perform data mining on said database.

51. Method of building a stochastic model of a data sequence, said data sequence comprising time related symbols selected from a finite symbol set, the method comprising:

receiving said data sequence,

expressing said symbols as a series of counters within nodes, each node having a counter for each symbol, each node having a position within said tree, said position expressing a symbol sequence and each counter indicating a number of its corresponding symbol which follows a symbol sequence of its respective node, and

reducing said tree to an irreducible set of conditional probabilities of relationships between symbols in said input data sequence, thereby to model said sequence.